

Supplementary Material

PREDICT: A method for inferring novel drug indications with application to personalized medicine

Assaf Gottlieb¹, Gideon Y. Stein^{2,3}, Eytan Ruppin^{1,2} & Roded Sharan¹

¹ *The Blavatnik School of Computer Science, Tel-Aviv University, Tel-Aviv, Israel 69978*

² *Sackler School of Medicine, Tel Aviv University, Tel Aviv, Israel 69978*

³ *Dept. of Internal Medicine "B", Beilinson Hospital, Rabin Medical Center, Petah-Tikva, Israel 49100*

Supplementary tables

Table S1. Drug Indications gold standard

Table is supplemented as an Excel file.

Table S2. Predicted repositioning of drug Indications

Table is supplemented as an Excel file.

Table S3. Predicted drug indications for experimental drugs

Table is supplemented as an Excel file.

Table S4. Drug predictions for Paget's disease of bone.

Score	Name
1	1-hydroxy-2-(3-pyridinyl)ethylidene bis-phosphonic acid
0.98	[1-hydroxy-3-(methyl-pentyl-amino)-1-phosphono-propyl]-phosphonic acid
0.88	n-{1-[5-(1-carbamoyl-2-mercapto-ethylcarbamoyl)-pentylcarbamoyl]-2-[4-(difluoro-phosphono-methyl)-phenyl]-ethyl}-3-{2-[4-(difluoro-phosphono-methyl)-phenyl]-acetylamino}-succinamic acid

0.77	7-(1,1-dioxo-1h-benzo[d]isothiazol-3-ylloxymethyl)-2-(oxalyl-amino)-4,7-dihydro-5h-thieno[2,3-c]pyran-3-carboxylic acid
0.75	[(4-{4-[4-(difluoro-phosphono-methyl)-phenyl]-butyl}-phenyl)-difluoro-methyl]-phosphonic acid

Table S5. Statistics of overlap between the signature-based predictions and drug indications that are under clinical trial.

	# of clinical trials associations		Predicted	
Phases	Total	Known	Coverage	P-value
All	1884*	456	16%	$1.4 \cdot 10^{-30}$
I	802	232	20%	$2.7 \cdot 10^{-24}$
II	1247	321	16%	$1.2 \cdot 10^{-20}$
III	860	327	28%	$5.6 \cdot 10^{-49}$
IV	389	197	22%	$2.8 \cdot 10^{-11}$
Unlisted	490	218	24%	$8.3 \cdot 10^{-18}$

* Unique associations, excluding redundancy between phases.

Table S6. Mapping between OMIM diseases and UMLS concepts

Table is supplemented as an Excel file.

Supplementary figures

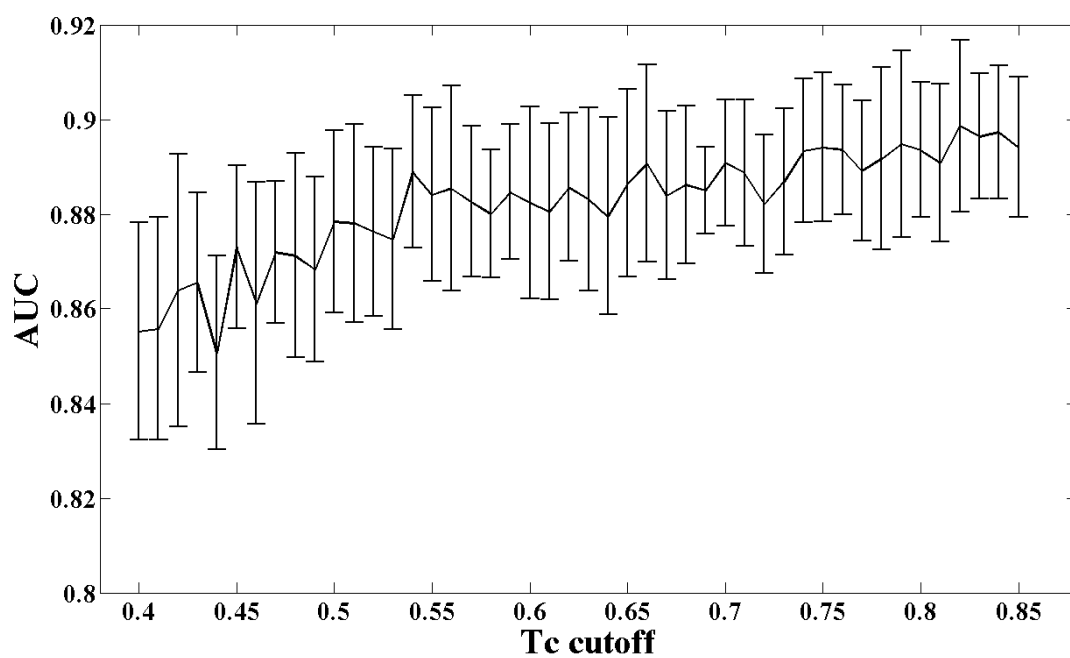


Figure S1. AUC scores as a function of the Tanimoto coefficient (Tc) cutoff.

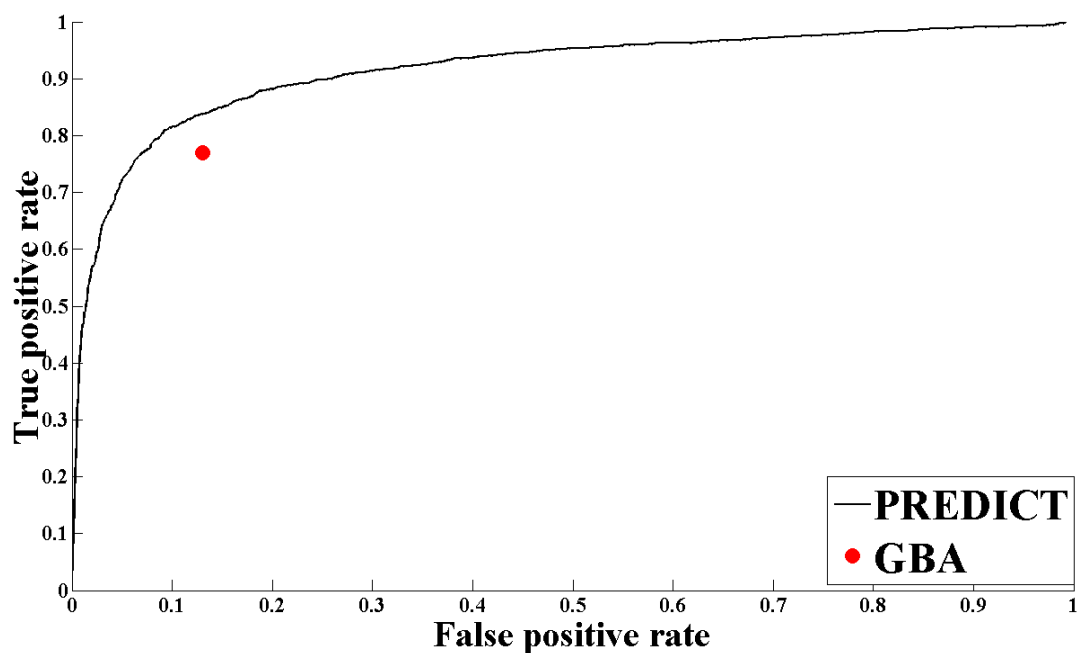


Figure S2. Receiver-Operating Characteristic (ROC) curve of our method and the FPR-TPR point of the GBA method.

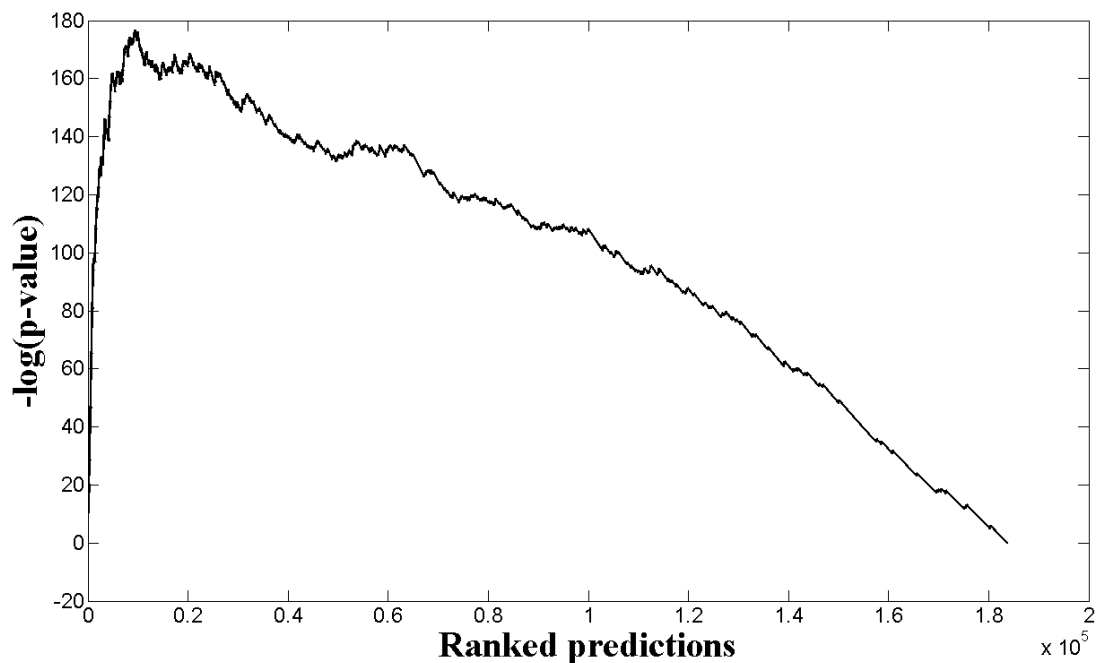


Figure S3. Hypergeometric p-values, obtained at different score thresholds, for the enrichment of low-confidence associations in our ranked predictions.

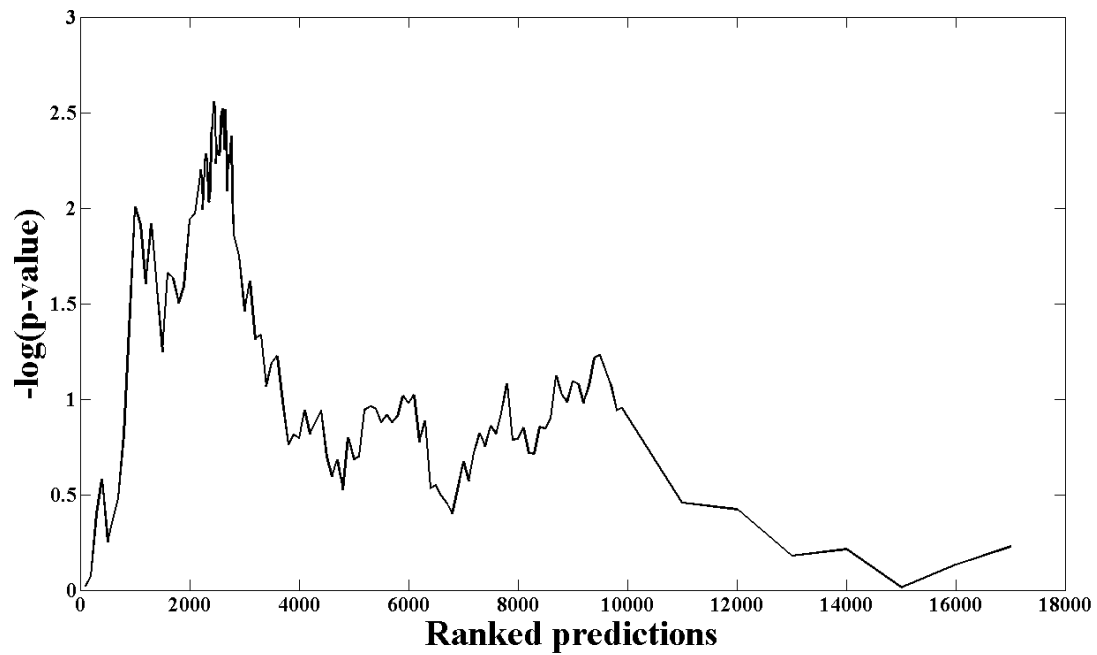


Figure S4. Hypergeometric p-values, obtained at different score thresholds, for the agreement between tissues associated with a drug target and those associated with the diseases it is predicted to treat.

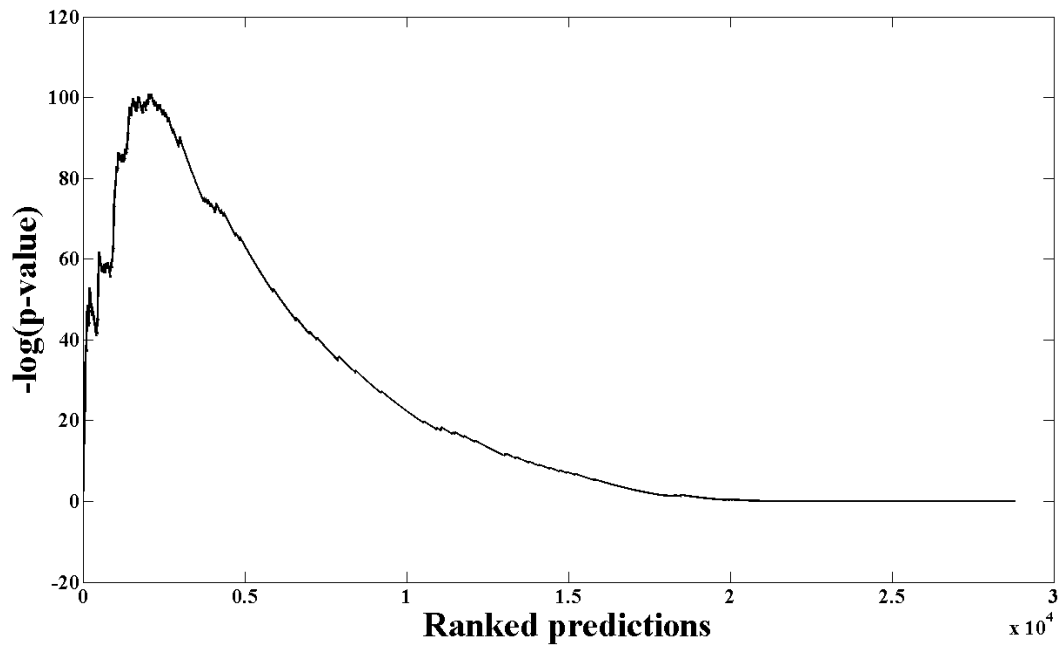


Figure S5. Hypergeometric p-values, obtained at different score thresholds, for the enrichment of low-confidence associations in our signature-based predictions.